

IVVI: Intelligent vehicle based on visual information

José María Armingol^{a,*}, Arturo de la Escalera^a, Cristina Hilario^a, Juan M. Collado^a,
Juan Pablo Carrasco^a, Marco Javier Flores^a, José Manuel Pastor^b, F^o José Rodríguez^a

^a *Intelligent Systems Laboratory, Department of Systems Engineering and Automation, Universidad Carlos III de Madrid, Spain*

^b *Department of Informatics Systems, Escuela Universitaria Politécnica de Cuenca Universidad de Castilla la Mancha, Spain*

Abstract

Human errors are the cause of most traffic accidents, with drivers' inattention and wrong driving decisions being the two main sources. These errors can be reduced, but not completely eliminated. That is why Advanced Driver Assistance Systems (ADAS) can reduce the number, danger and severity of traffic accidents. Several ADAS, which nowadays are being researched for Intelligent vehicles, are based on Artificial Intelligence and Robotics technologies. In this article a research platform for the implementation of systems based on computer vision is presented, and different visual perception modules useful for some ADAS such as Line Keeping System, Adaptive Cruise Control, Pedestrian Protector, or Speed Supervisor, are described.

Keywords: Driver assistance systems; Intelligent vehicles; Intelligent transportation systems; Computer vision

1. Introduction

It is well-known that human errors are the cause of most traffic accidents. The two main errors are drivers' inattention and wrong driving decisions. Governments are trying to reduce the said accidents with infrastructure improvement and educational campaigns, but they cannot be completely eliminated due to the human factor. That is why ADAS can reduce the number, danger and severity of traffic accidents. Several ADAS, which nowadays are being researched for Intelligent vehicles, are based on Artificial Intelligence and Robotics technologies.

Computer vision plays an important role [1]. Drivers receive a good deal of visual information, of different kinds while driving a car. It can be stated that this information is related to two main goals: safe and successful driving. These goals are achieved through two driver's tasks: piloting and navigation. The first task takes into account visual information, in order to avoid obstacles while driving the vehicle in the correct lane

and at the right speed: position and trajectory of other vehicles, pedestrians' movements, road lanes and vehicle status.

Other visual information such as landmark detection and recognition are related to the success of the driving, and to the arrival at the destination, which is the goal of the driver's navigation task. ADAS are designed to help (or substitute) human drivers in these two tasks. Thus, there are examples of road lane detection and following [2,3], obstacle recognition and avoidance in the vehicle's path like: vehicles [2,4] or pedestrians [4,5], or other elements such as traffic lights and marks on roads [6]. Traffic or road signs and route guidance signs are located in streets, roads and motorways to help drivers. The first ones are designed to help the piloting task, providing information on the maximum or minimum speed allowed, shape of the road, prohibited manoeuvres, etc. Route guidance signs help in the navigation task, providing information on destination, junctions, exits, distances, etc. ADAS with the automatic ability of extracting and identifying these signs would help human drivers a great deal, making the navigation task easier and allowing the driver to concentrate on driving of the vehicle.

For these reasons, ADAS based on Computer vision can help human drivers in many ways, for example:

- *Lane Keeping/Departure Warning System.* The vehicle's relative position with the lines that mark the limits of the

* Corresponding address: Escuela Politécnica Superior, Universidad Carlos III de Madrid, C./ Butarque 15, 28911 Leganes, Spain. Tel.: +34 916245926; fax: +34 916249430.

E-mail address: armingol@ing.uc3m.es (J.M. Armingol).

URL: <http://www.uc3m.es/islab> (Intelligent Systems Lab).

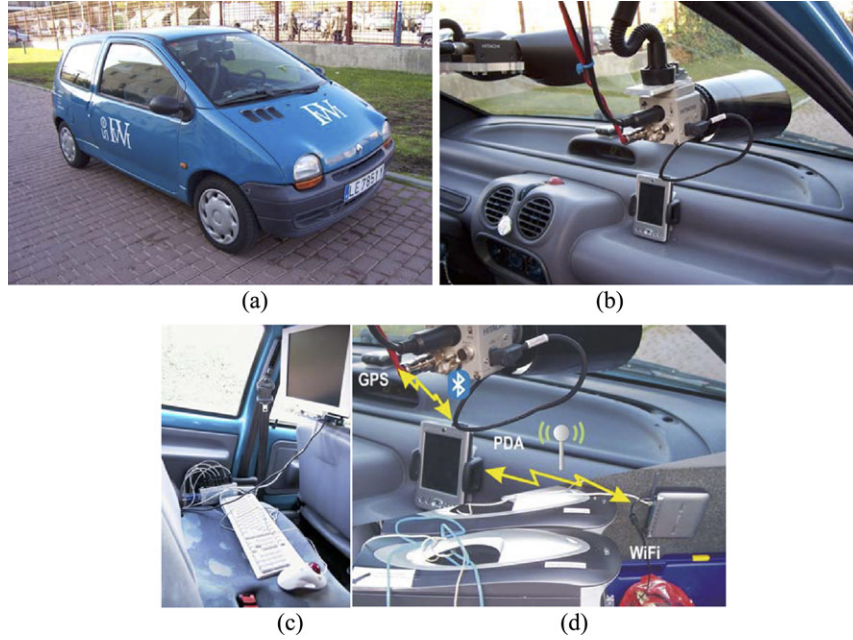


Fig. 1. IVVI vehicle. (a) General view (b) stereo system and colour camera (c) operator workplace (d) GPS and communication system.

lane is found through the analysis of the images grabbed by the camera. The alarm is set off when the trajectory of the vehicle differs from the ideal one (Lane Departure Warning System) or the steering wheel offers resistance to move proportionally to the distance from the lane limits (Lane Keeping System).

- **Pedestrian Protector.** Pedestrians are the weakest and most unprotected elements among all the objects in motion on streets and roads. The pedestrian detection system has to warn the driver if there is a possible collision. To do so, the system not only has to detect all the pedestrians who are around the vehicle, but in addition, has to analyse their activity and movement in order to determine and localize those cases where there would be a conflict.
- **Adaptive Cruise Control.** The system recognizes preceding vehicles and calculates their speed and trajectory. It determines the vehicle maximum speed in order to maintain a safe distance from the other vehicles. This module has as its main difficulty the fact that objects in the image can have great diversity in appearance. This is due to the existence of several vehicle models and the camera's perspective changes.
- **Traffic sign perception.** Perception of traffic signs is very useful to check if the vehicle's speed is the appropriate one and to warn the driver of possible dangerous situations which can appear, etc. Automatic sign recognition is complicated because of the presence of other objects with the same colours, or due to object occlusions or to changeable lighting condition.

The overview of the paper is: Section 2 focuses on the IVVI (Intelligent vehicle based on Visual Information) research platform: its goal and the hardware elements. Section 3 describes the module for detection and classification of road lanes (continuous, discontinuous or merge), and finally the

road type (one-way, two-way or highway). Section 4 is for pedestrian detection where the active contour paradigm and stereo vision are used. Section 5 is devoted to vehicle detection. A geometric model is presented and a genetic algorithm finds its optimum parameter values. A multi-resolution approach is used to speed up the algorithm and work in realtime. Finally Section 6 describes an algorithm for traffic sign detection and classification. Based on their colour and shape, the signs are detected through normalized correlation.

2. The IVVI experimental vehicle

IVVI (Fig. 1(a)) is a research vehicle for the implementation of systems based on computer vision, with the goal of building ADAS. The purpose of the IVVI platform is to test computer vision algorithms under real conditions, and at this moment four sensing capabilities are being researched for road lanes, pedestrians, vehicles and traffic signs. They can be taken as inputs for some ADAS like Lane Keeping System, Adaptive Cruise Control, Pedestrian Protector, and Traffic Sign Recognition. All of the designed ADAS work in urban and road environments, providing realtime information to the driver.

Research results are being implemented in an actual vehicle (presently, a Renault Twingo). There is a DC/AC power converter connected to the vehicle's battery. Through it, the electrical power needed for the computers and cameras is obtained. There are two PCs (Fig. 1(d)) in the vehicle's boot which are used for processing the images captured by the cameras. There is an electronic multiplexer for the video, mouse and keyboard signals so that the human operator, placed on the back seat (Fig. 1(c)), is able to work with two systems simultaneously. There is a binocular stereo vision system (Fig. 1(b)). As the cameras are progressively scanning, they can capture images on movement and avoid the problems associated

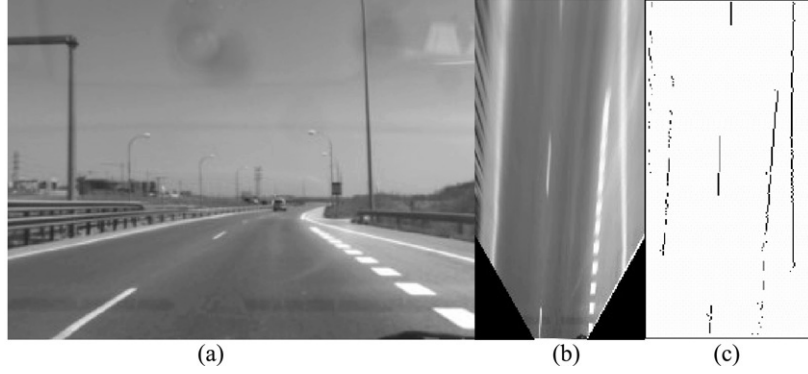


Fig. 2. (a) On-board camera view; (b) remapped image; (c) thresholded image.

with interlace video. There is a colour CCD camera for the detection of traffic signs and other vertical signs (Fig. 1(b)). The vehicle's speed, position and course are obtained by a GPS connected to a PDA by Bluetooth. The PDA broadcasts the information to the PCs through a WiFi router (Fig. 1(d)).

As one of the computers receives the information from the stereo system, it stores the results of the lane, vehicles and pedestrian detection; the second computer analyses the information of the detected traffic signs. This way, one computer prevents car crashing (lane departure, collision against an obstacle), while the second attends to traffic offences. In order to give any information to the driver both computers are connected to the loudspeakers of the vehicle and different audio warning messages are broadcasted. In case of conflict, a possible collision has the highest priority followed by lane departure and traffic infringements.

The detection of road lanes is useful for the vehicle and pedestrian detection modules. From the number and location of the detected lanes some geometric restrictions can be applied to the position, width and height of a vehicle in the image. The possible area in the image where a vehicle can appear is delimited and the scale of meters per pixels for every road in the image is known. In a similar way, pedestrians can appear every where in the image as they can cross the road or walk on the pavement. A stereo vision system is used for this detection; and the range information can be applied in conjunction with the geometric restrictions mentioned before. There is no link between the road and traffic sign module for the detection process. A previous detection of the road could be useful to the traffic sign detection module in two ways: to discharge some region in the image as traffic signs are not within road limits, and in highways traffic signs are near the hard shoulder. In the implementation described in this article these two considerations have not been taken into account: the first one is not taken into account because colour is a basic feature to detect the traffic signs and as roads are nearly colourless (grey scale), they are easily discharged by the colour analysis step. The second one is not considered because in urban environments there are no regulations about traffic sign placement and cannot be applied.

In the following sections, these perception modules will be described in more detail.

3. Detection and classification of road lanes

The purpose of the Road Detection and Interpretation capability is to calculate the lane position automatically with respect to the car, the lane boundary type (continuous, discontinuous or merge), and finally the road type (one-way, two-way or highway). The latter can only be found from the number of lanes and the line types.

Lane detection has been researched for more than two decades. Some striking results have already been demonstrated under real conditions [2,3,7], especially in the area of automatic driving. Nevertheless, at the present time, research is moving towards driver assistance systems because of the legal and psychological difficulties of an automatic system. Today, even though many car manufacturers have expressed the intention to include vision-based ADAS in the short term, there are still very few commercial systems available for commercial cars. It can be said that though many problems have been already solved, there is still a need to increase the robustness of these systems so that they can be useful to drivers.

3.1. Road line detection

The image analysis can be done in two reference systems. That is, from the car-view image [8] (Fig. 2(a)) and from a bird's eye-view after a perspective transformation [9], assuming that the world is flat (Fig. 2(b)). The latter is the option chosen in this work, which has the advantage of facilitating the detection of the marked lines and the integration of temporal information, but presents calibration problems. If the extrinsic calibration parameters of the vision system (its position and orientation in world coordinates), are not well-calculated, the flat road assumption is violated, and the bird's eye-view image will show converging or diverging lines, instead of parallel ones. This leads to a bad calculation of the lane position and orientation. To overcome these problems, an auto-calibration algorithm based on evolutionary techniques has been tested on the IVVI vehicle [10]. The lane detection algorithm is designed to work in painted roads, outside cities; therefore, in this case the relevant characteristics are the edges of the objects in the image. Road lines are usually modelled as bright bands over a darker background. As the lane curvature is small in the nearby region of the road, these lines are mainly vertical in the bird's

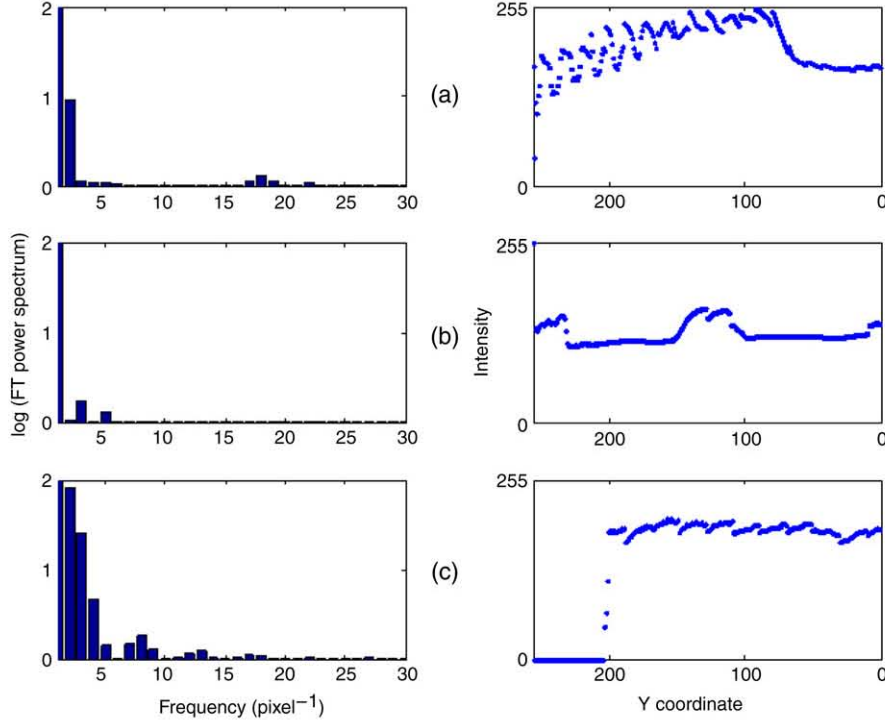


Fig. 3. Road lines classification (right) line profile extracted from the intensity image (Fig. 2(b)) (left). Power spectrum of the Fourier analysis (a) merge line; (b) discontinuous line; (c) continuous line.

eye-view image of the road. Therefore, the search for road mark candidates consists of looking for dark–bright–dark transitions in the horizontal direction.

The borders of the image are extracted with a spatial filter based on the ideas of the Canny border extractor. The Canny filter offers a good signal–noise ratio, compared to other border extractors. This is important because the border image will be used in the frequency analysis that identifies the road line type; thus, this image should be as noise clean as possible. In this article, a straight line model is extracted. Although this model gives a poor description of the road shape, it has several advantages. Straight lines can be robustly and quickly extracted with the Hough Transform, a technique that can hardly be applied to more complex models in realtime. It also facilitates greatly the process of auto-calibration, and line type classification. In addition, this model fits quite well to the road lines in the nearby region of the road.

3.2. Road lines classification

The extracted lines are classified into the different types that are found on roads. The main difficulty of this task is the lack of international standardization regarding the length and frequency of the white stripes in dashed lines. However, most roads have three basic line types, namely: continuous, discontinuous and merge.

The intensity line profile for each detected line (right column of Fig. 3) is not a good data to feed into the frequency analysis, because its appearance changes substantially with the environmental conditions. Furthermore, the resolution of the bird’s eye-view in the distance is poor. This is a drawback with

the merge lines, which appear blurred far ahead and can even look like continuous ones (Fig. 3(a)). Furthermore, the power spectrum (left column of Fig. 3) presents a tiny peak at the specific frequency.

It is more robust to obtain the line profile from the thresholded image given by the preprocessing step (Fig. 2(c)), which is shown on the left side of Fig. 4.

Again, the right side of the figure depicts the power spectrum of the Fast Fourier Transform applied to the line profile vector. The results show that a clear sharp peak appears in the Fourier Transform power spectrum when the line is dashed, and that the value of the frequency associated with that peak gives the line type (discontinuous or merge). These peaks are depicted on the left side of Fig. 4(a) and (b) with arrows pointing at them. No significant peaks are present when the line is continuous (Fig. 4(c)). It can be seen that, now, the peaks are more pronounced and much easier to detect. It has been heuristically found that only the first 21 frequencies are significant in this analysis.

The algorithm works on board the IVVI vehicle in realtime and has been tested under real conditions. The lane detection is robust to shadows and partial occlusions, as showed in Fig. 5. An audio warning is sent by the vehicle loudspeakers if the vehicle is about to cross a continuous line or a discontinuous line if the indicator is not active. The intersection of the detected lines is the vanishing point. The pitch of the camera can be obtained and the pedestrian detection module will use this information for the stereo base segmentation. The vehicle detection module receives information on the position and orientation of the detected lanes.

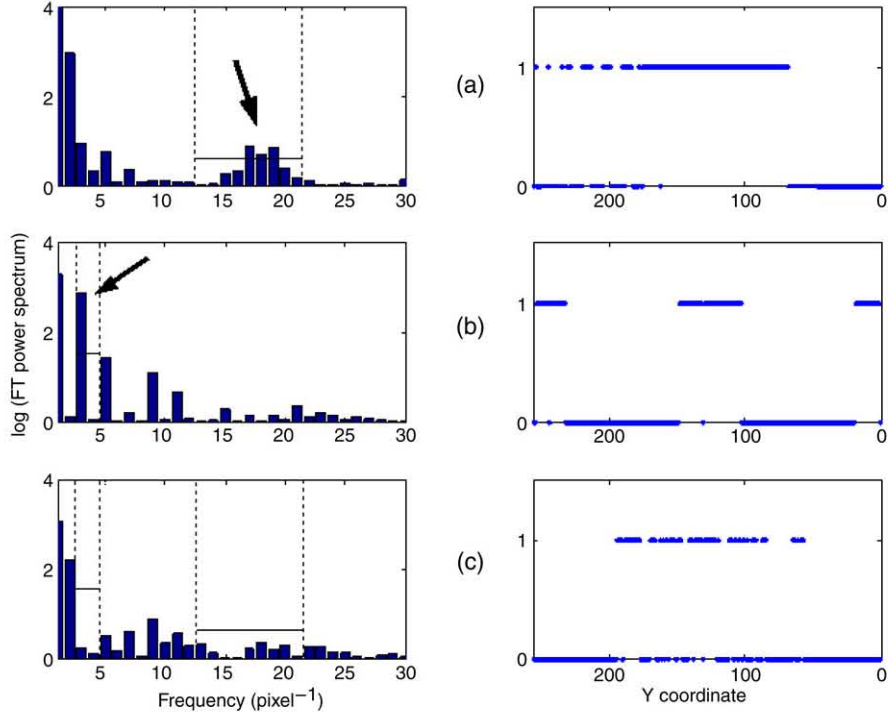


Fig. 4. Fourier analysis for the line profile extracted from the mark candidates' image (Fig. 2(c)). (a) Merge line; (b) discontinuous line; (c) continuous line.

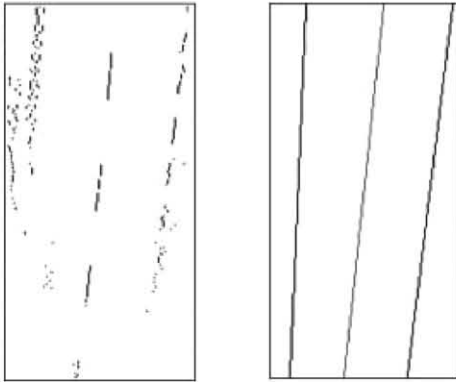


Fig. 5. Example of road lines detection and classification.

4. Pedestrian detection

An analysis of European traffic accident shows a very large number of collisions between pedestrians and vehicles. Every year in the European Union, more than 200,000 pedestrians

are injured, with over 6000 killed. These collisions account for about 12% of total road accidents and about 15% of the total number of road deaths.

In this article a pedestrian detection method based on an active contour model is presented. Stereo vision is used to initialize the model and to improve the accuracy of the segmentation. A priori knowledge of the object shape is exploited to control the deformation of the model. Unlike classical formulation, in this proposal image energies are spread along a large distance, so the contour can be attracted not only by near features.

4.1. Active contour models

Image segmentation based on active contours has received considerable attention in the past 20 years. Its success resides in its capability to integrate physical and topological knowledge into the segmentation process. Therefore, deformable models are able to extract contours, even in the presence of gaps or occlusions. Its dynamic behaviour makes this approach adequate for the detection and tracking of non-rigid objects. In this section, an active contour model is proposed to detect and track people from images taken by an on-board vision system, performing contour extraction in sequential frames. Our purpose is to develop an algorithm to protect pedestrians as part of an ADAS [11]. Our method differs from other approaches based on more rigid models like in [12], or methods that require a huge pedestrian appearance database [13] or even when they lack a model [14].

In the system developed, stereo vision is used to generate a disparity map of the scene (Fig. 6(b)). As the pedestrians can appear in the scene at very diverse distances, the use of range information allows filtering of the images based on distance

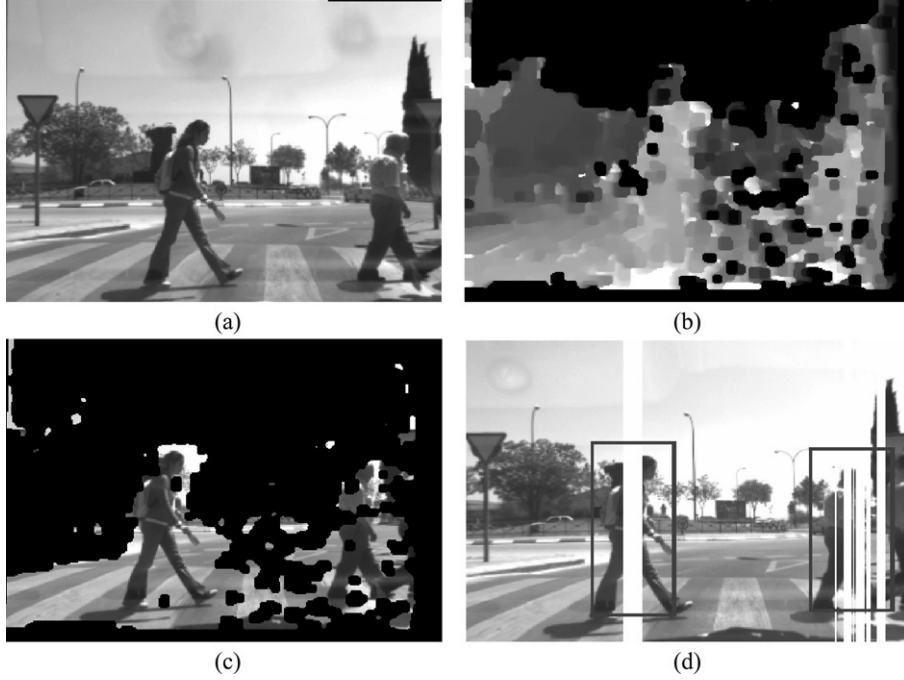


Fig. 6. (a) Right image taken by the stereo camera. (b) Disparity map (c) and the filtered image based on distance measures. (d) Both vertical symmetries and snakes initialization region enclosed by bounding boxes are shown.

measures. Therefore, regions that are not at the desired distance are eliminated (Fig. 6(c)), performing subsequent calculations only on the filtered areas. Hence, two advantages are obtained; Firstly, the algorithm is less time consuming as a complete scan of the image is avoided. Secondly, the task of initializing the snakes is eased because only the filtered area is considered. Pedestrians present great diversity of appearance but all have in common a strong vertical symmetry. Because of that, regions with high vertical symmetry are potential candidates for an active contour initialization. With the aim of finding these regions the vertical gradient component of the filtered image is found and only pixels with high response are taken. Then, pairs of pixels on the same line vote for central pixels as their symmetry axis. An active contour is initialized in a symmetry axis, if the number of pixels that vote for that axis is over a given threshold (Fig. 6(d)).

4.2. Contour evolution

Once the snake has been initialized it deforms by external and internal forces. For the external forces, a new potential field which smoothly extends over a long distance is proposed. So, the snake is affected not only by surrounding features. The fact that pedestrians have a strong vertical symmetry is exploited to construct the potential field. Therefore, a distance map of the symmetry axes obtained from that stage is constructed. In order to avoid the snake shrinking to the axis, its movement is allowed until it reaches a vertical edge. Besides this potential field, the image gradient and distances to vertical borders are also considered (Fig. 7).

The internal energy of the contour represents its shape as a combination of both membrane and thin-plate behaviours. The first term in the formulae controls the discretization of the

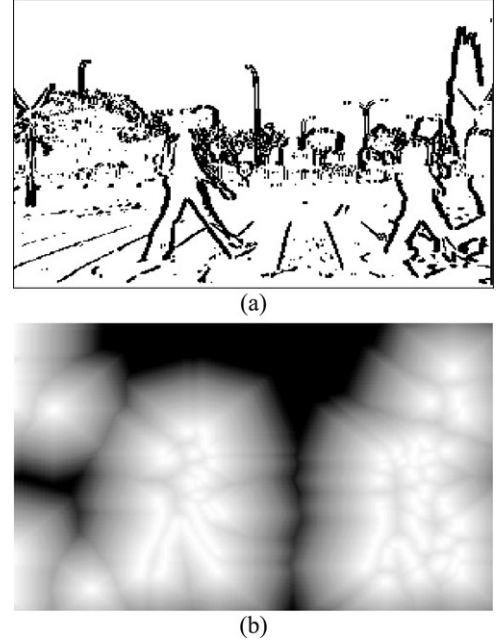


Fig. 7. (a) Vertical gradient and (b) its distance map.

contour, moving the points along its tangential component in order to maintain the same spacing between them. The second term determines the extent to which the snake can bend. These elastic forces are used to regularize the contour and hence to avoid strange effects.

$$E_{\text{int}}^*(v(s)) = \alpha(s) \left(\overline{\text{dist}} - \left| \frac{dv}{ds} \right|^2 \right) + \beta(s) \left| \frac{d^2v}{ds^2} \right|^2. \quad (1)$$

In order to avoid shrinkage, a new internal force is included to control the shape of the deformable model. This regularizing

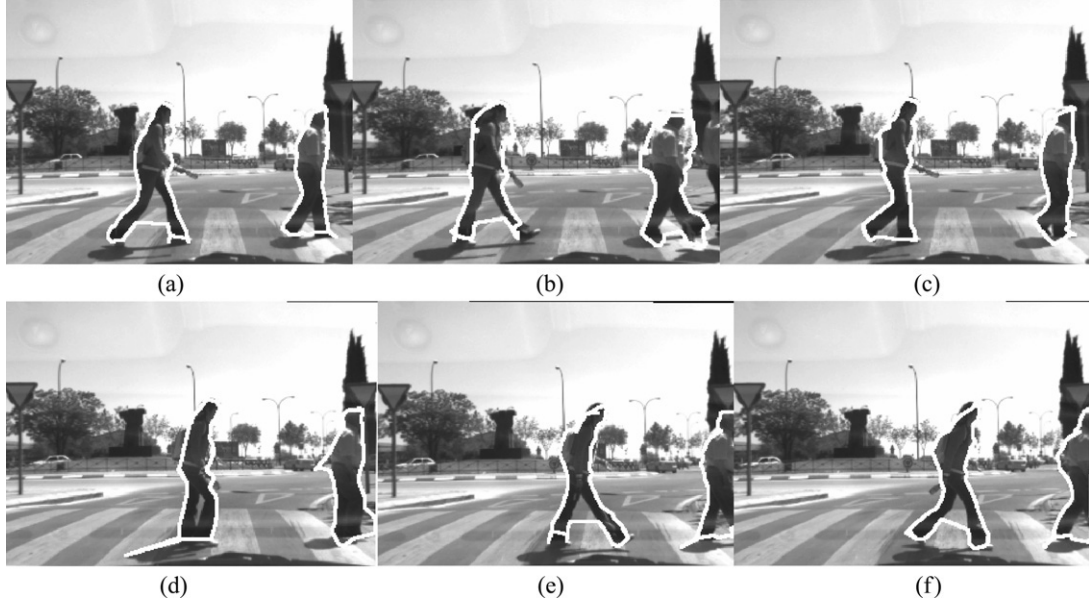


Fig. 8. From (a) to (f) a complete sequence of frames are shown.

force prevents the shrinking effect of the snake, as it is based on higher degrees of smoothness than the membrane and the thin-plate energy terms, which are based on the first and second derivatives respectively.

$$E_{\text{int}}(v(s)) = E_{\text{int}}^*(v(s)) + \theta(s) \left| \frac{d^4 v}{ds^4} \right|^2. \quad (2)$$

The deformable model proposed extends the greedy algorithm of Williams and Shah [15] as it is a stable, fast and flexible optimisation technique. This approach is adequate for non-rigid objects' detection and tracking, performing contour extraction in sequential frames. Once the snake is initialized on an object contour in the first frame, it will automatically track the contour from frame-to-frame. This method requires small deformation and movement of an object between frames.

Some points in the snake are still prone to errors, such as getting trapped in the shadow of the pedestrian (Fig. 8(d)). In addition, if the external forces are not strong enough, the snake tends to shrink (Fig. 8(b) and (e)).

If the speed of the vehicle is above a threshold and a pedestrian is detected in the central columns of the image an audio warning beep is sent to the driver to advice him/her of the presence of an obstacle.

5. Vehicle detection

Intelligent vehicles have to detect and track other vehicles. At present, commercial equipment is based on distance sensors like radar or laser. These sensors have the advantage of giving a direct distance measurement and, above all, they are able to work under bad weather conditions. Their main drawback is the field of vision, which is very narrow, so they can only detect the vehicle in front of the sensor. If the vehicle is overtaken, there is a step input to the system and the response can be unstable. One alternative or complementary sensor is vision.

Although it is not able to work under bad weather conditions and its information is more difficult to process, it gives a richer description of the environment that surrounds the vehicle [16]. Furthermore, many of the current traffic accidents happen in good weather and are due to human errors.

In this article vehicle detection is obtained through a top-down approach. A geometric model of the vehicle is defined where its energy function includes information about the shape and symmetry of the vehicle and the shadow it produces. A genetic algorithm finds the optimum parameter values. As the algorithm receives information from a road detection module (Section 3) some geometric restrictions can be applied. A multi-resolution approach is used to speed up the algorithm and work in realtime.

5.1. Geometrical model of a vehicle

Due to shadows, occlusions, weather conditions, etc, the model has to incorporate as much information as possible. In this article, a vehicle is defined by seven parameters (Fig. 9(a)): Position (x, y), width and height of the vehicle, windshield position, bumper position and roof angle. The previous detection of the road limits helps in the vehicle detection step because the searched area is smaller (Fig. 9(b)).

The energy function considers the following three factors: symmetry, shape and the vehicle shadow (Fig. 10).

The symmetry of the vertical and horizontal edges is considered. For this reason, the vertical and horizontal gradient components of the image are found (Fig. 10(b) and (c)). Only the pixels with a high response in one of the components and low in the other are taken into account. Then, the pairs of pixels in the same line vote for the central pixel as their symmetry axis, following the formulae:

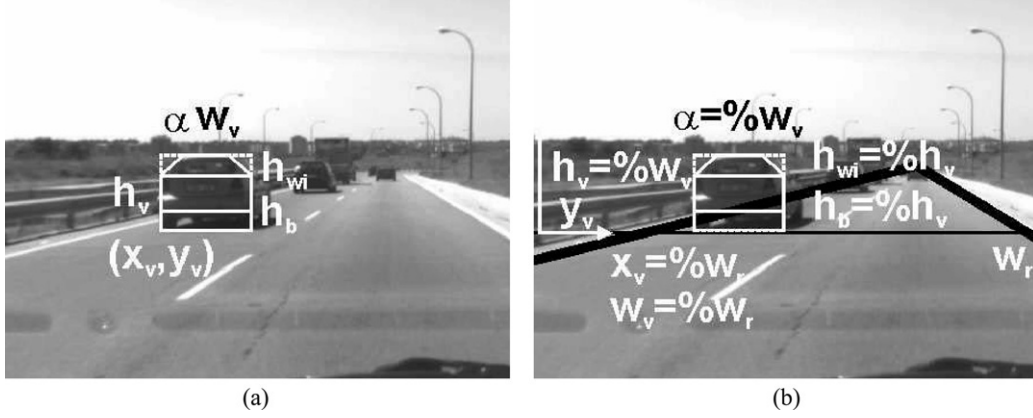


Fig. 9. Geometrical model of a vehicle. (a) Parameters (b) the values of these parameters are constrained by the detection of the road.

$$\begin{aligned}
 &\text{If } \begin{cases} D_{\min} < |x_i - x_j| < D_{\max} \\ y_i = y_j \\ G_v(x_i, y_i) > T_v \\ G_v(x_j, y_j) > T_v \end{cases} \quad \text{then } +S_v(k)k = \frac{i+j}{2} \\
 &\text{If } \begin{cases} D_{\min} < |x_i - x_j| < D_{\max} \\ y_i = y_j \\ G_h(x_i, y_i) > T_h \\ G_h(x_j, y_j) > T_h \end{cases} \\
 &\text{then } +S_h(k)k = \frac{i+j}{2}. \quad (3)
 \end{aligned}$$

Where D_{\min} and D_{\max} are the minimum and maximum width of the searched vehicle in the image, G_h and G_v the vertical and horizontal components of the gradient, T_h and T_v are the thresholds, and S_v and S_h are the accumulator arrays for the symmetry axis. For a specific model, the value of the symmetry energy is:

$$\begin{aligned}
 E_{SV} &= \frac{Sv(x+w/2)}{h} & E_{SH} &= \frac{Sh(x+w/2)}{4w} \\
 E_{Sim} &= \frac{1}{2}(E_{SiV} + E_{SHm}) \quad (4)
 \end{aligned}$$

where h is the height of the vehicle model, w is its width, E_{SV} the symmetry energy of the vertical edges, E_{SH} the symmetry energy of the horizontal edges and E_{Sim} the global symmetry energy.

Shape is defined by two energy terms: one based on the gradient (Fig. 10(b)) and the other one based on the distance from the edges, found before for the symmetry energy (Fig. 10(d)). For a particular model with height h , width w , position (x, y) , windshield position t , and bumper position m , the vertical gradient energy, E_{GV} , and horizontal gradient energy, E_{GH} , are:

$$\begin{aligned}
 E_{GV} &= \frac{1}{2h} \left(\sum_{j=y}^{y+h} Gv(x, j) + \sum_{j=y}^{y+h} Gv(x+w, j) \right) \\
 E_{GH} &= \frac{1}{4w} \left(\sum_{i=x}^{x+w} Gh(i, y) + \sum_{i=x}^{x+w} Gh(i, y+t) \right. \\
 &\quad \left. + \sum_{i=x}^{x+w} Gh(i, y+m) + \sum_{i=x}^{x+w} Gh(i, y+h) \right). \quad (5)
 \end{aligned}$$

And the global gradient energy E_G :

$$E_G = \frac{(E_{GV} + E_{GH})}{2}. \quad (6)$$

A distance image is obtained where each pixel shows the distance from the nearest edge. In order to emphasize the pixels that are near the edges, the following lookup table is applied

$$Lut(D) = \begin{cases} 255(1 - \text{sqrt}(D/D_{\max})) & 0 < D < D_{\max} \\ 0 & D > D_{\max} \end{cases}. \quad (7)$$

From that image, a distance from vertical edge energy, D_{GV} , and horizontal edge energy, D_{GH} , are calculated, where D_G is the global distance energy.

$$\begin{aligned}
 D_{GV} &= \frac{1}{2h} \left(\sum_{j=y}^{y+h} Dv(x, j) + \sum_{j=y}^{y+h} Dv(x+w, j) \right) \\
 D_{GH} &= \frac{1}{4w} \left(\sum_{i=x}^{x+w} Dh(i, y) + \sum_{i=x}^{x+w} Dh(i, y+t) \right. \\
 &\quad \left. + \sum_{i=x}^{x+w} Dh(i, y+m) + \sum_{i=x}^{x+w} Dh(i, y+h) \right) \quad (8) \\
 D_G &= \frac{(D_{GV} + D_{GH})}{2}.
 \end{aligned}$$

The shadow energy, E_{SOM} , of a vehicle with height h , width w , position (x, y) , and bumper position m , is defined by the average level of grey in the lower part of the model:

$$E_{SOM} = 1 - \frac{\sum_{i=x}^{x+w} \sum_{j=y+m}^{y+h} I(i, j)}{(h-m)w}. \quad (9)$$

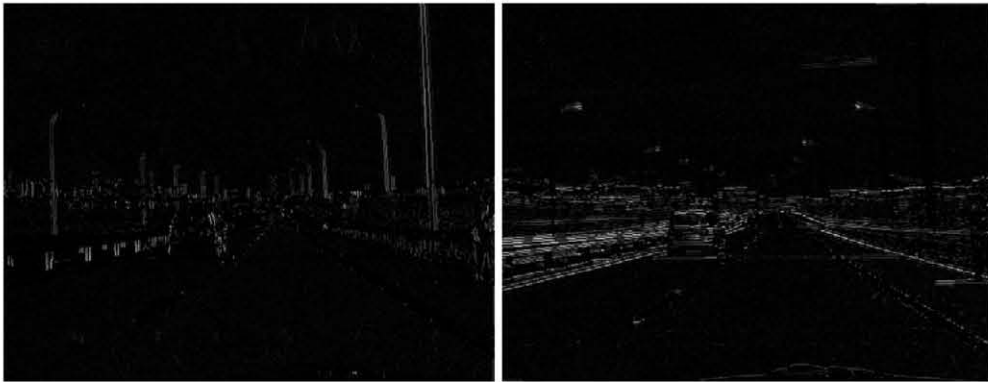
If it is black, the energy value is one and zero if it is white. The final energy, E , is:

$$\begin{aligned}
 E(Z) &= -(k_A E_{Sim}(Z) + k_B E_G(Z) \\
 &\quad + k_C E_D(Z) + k_D E_{SOM}(Z)) \quad (10)
 \end{aligned}$$

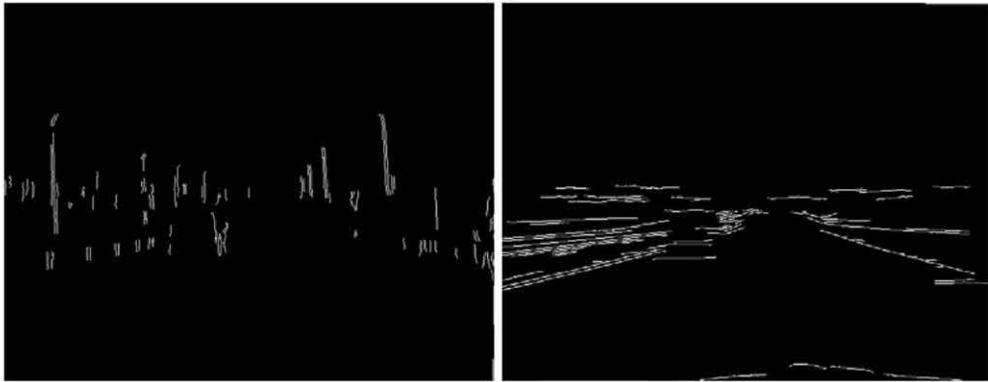
where k_A , k_B , k_C and k_D allow a weighted sum of the energy terms.



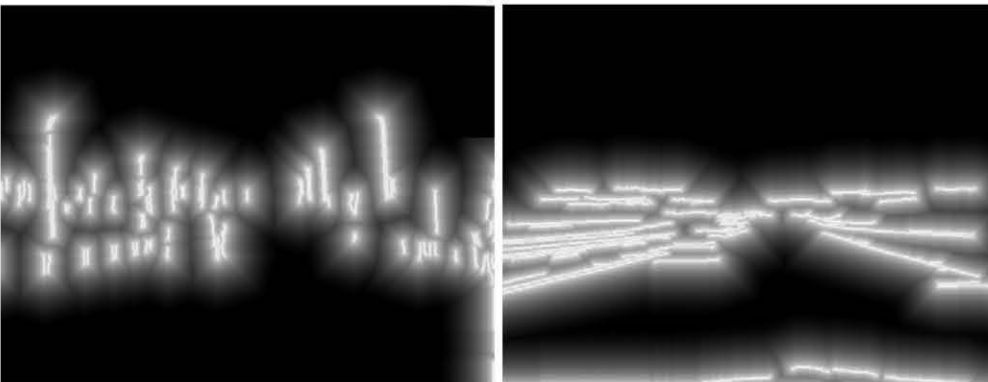
(a)



(b)



(c)



(d)

Fig. 10. Image processing (a) image (b) vertical and horizontal gradients (c) vertical and horizontal edges (d) vertical and horizontal distances.

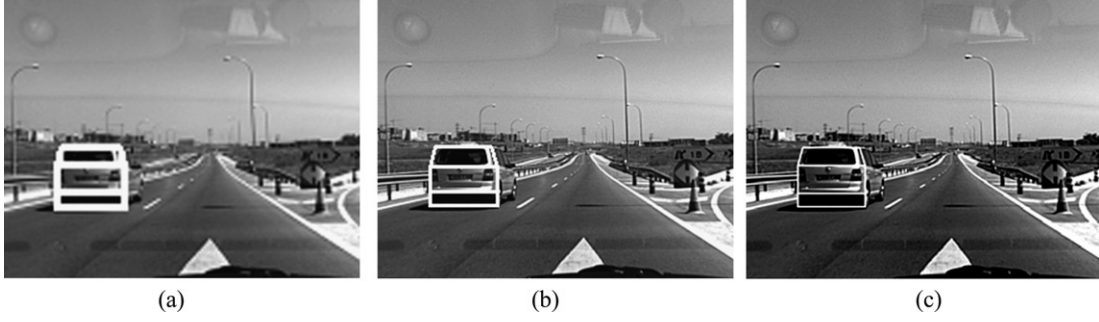


Fig. 11. Multi-resolution detection at (a) 160×120 (b) 320×240 (c) 640×480 pixels.

5.2. Likelihood probability density function

The estimate of a given deformation Z for the image I , $P(I|Z)$, follows a Gibbs distribution [17]:

$$P(I/Z) = \frac{1}{K} \exp -E(Z) \quad (11)$$

where K is the normalizing constant. The detection problem is the search of the Maximum A Posteriori (MAP) estimation of Z .

$$Z_{\text{MAP}} \in \arg \max_Z P(I/Z) \in \arg \min_Z E(Z). \quad (12)$$

The energy function is minimal when the deformed model exactly matches with the one presented in the image.

5.3. Search algorithm

Search algorithms have to find a balance between two opposite tasks: exploration of the complete search space and the exploitation of certain zones.

With exploration, the search space is covered looking for promising areas in which a more detailed search has to be done; that is the exploitation task, where the best solution is looked for in a zone known as suitable. The risk is being trapped in a local maximum or minimum. Hashing methods are the extreme case of exploration, where gradient-based methods (hillclimbing) are the extreme for exploitation. Genetic algorithms (GAs) [18] do a parallel search in several directions following an optimisation process, which imitates natural selection and evolution. To accomplish this task, there are a set of possible solutions (the individuals) that exchange information depending on the fitness of the result in the search for the global maximum. GAs robustness relies on their ability to reach a global maximum surrounded by local ones.

The detection of the vehicle is done for multiple resolutions. A Gaussian pyramid is built, with dimensions: 160×120 , 320×240 , and 640×480 pixels. The information about the detection of lower levels is passed to upper levels (Fig. 11). Working with a multi-resolution approach has the main advantage of working with the best resolution for every circumstance.

The information received from the lane detection module is useful, not only for limiting the area where the vehicle can be found, but also for knowing if the detected vehicle is in the same lane as ours. If so, a beep warning is sent to the driver.

6. Traffic sign detection and classification

This system detects and recognizes traffic signs and the vehicle's speed through the GPS and communication system as shown in Fig. 1(b) and explained in Section 2. The driver is warned when the speed is greater than the road limit speed or, within cities, there is a warning sign like a roundabout or pedestrian crossing, and when the speed is too high; the warning sound is through the vehicle's loudspeakers.

As colour is an important feature of traffic signs, it is used for the detection step. There are two main colour spaces used: RGB [19] and HSI [20], although grey levels are also used [21]. As the sign's shape can have different appearances in the image some techniques has been used to speed up the algorithm like Genetic Algorithms [22], Hierarchical Models [13] or Normalized Correlation [23]. The proposed method, first detects the signs through normalized correlation of some models of circular and triangular signs in two images; one instance where the colour red (for warning and prohibition signs) and another where the colour grey, that is achromatic, has been enhanced for end-of-prohibition signs. Once there is a sign, the classification is performed by normalizing the detected area through a normalized correlation with a database of 96 existing traffic signs.

6.1. Colour enhanced

Two studies have been carried out, one for the colour red and another for the colour grey (achromatic). For the first one the RGB and the HSI cases have been tested. Changes in weather condition strongly affect the RGB. Because of that, the enhanced red colour image is normalized (Fig. 12(b)). It can be seen that the sign can be detected on a foggy morning but a false positive also appears. The case of analyzing the HSI image is shown in Fig. 12(b), where no false positive appears, and for the general case gives better results. End-of-limit signs are grey (Fig. 12(d)). This colour is not well-defined in the HSI space because of its poor saturation value. In the RGB space the colour grey is the diagonal that joins the origin of the cube, black, with the maximum value, white: all these colours have the same value for the three components. After analysing the values for some signs a distance threshold from this line has been found. If the distance is below this value the pixel is achromatic (Fig. 12(e)).

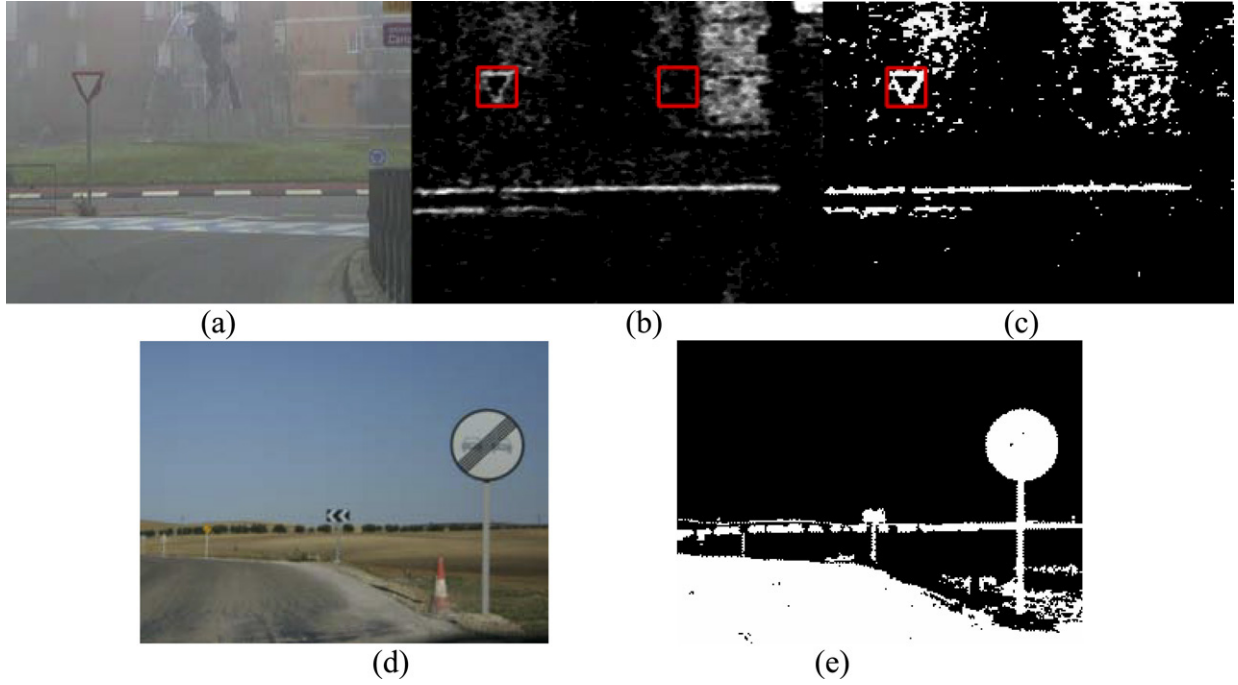


Fig. 12. Colour analysis. (a) An image taken during a foggy morning (b) the red colour enhanced using RGB (c) using HSI (d) image with an end-of-limit sign (e) achromatic pixels enhanced. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

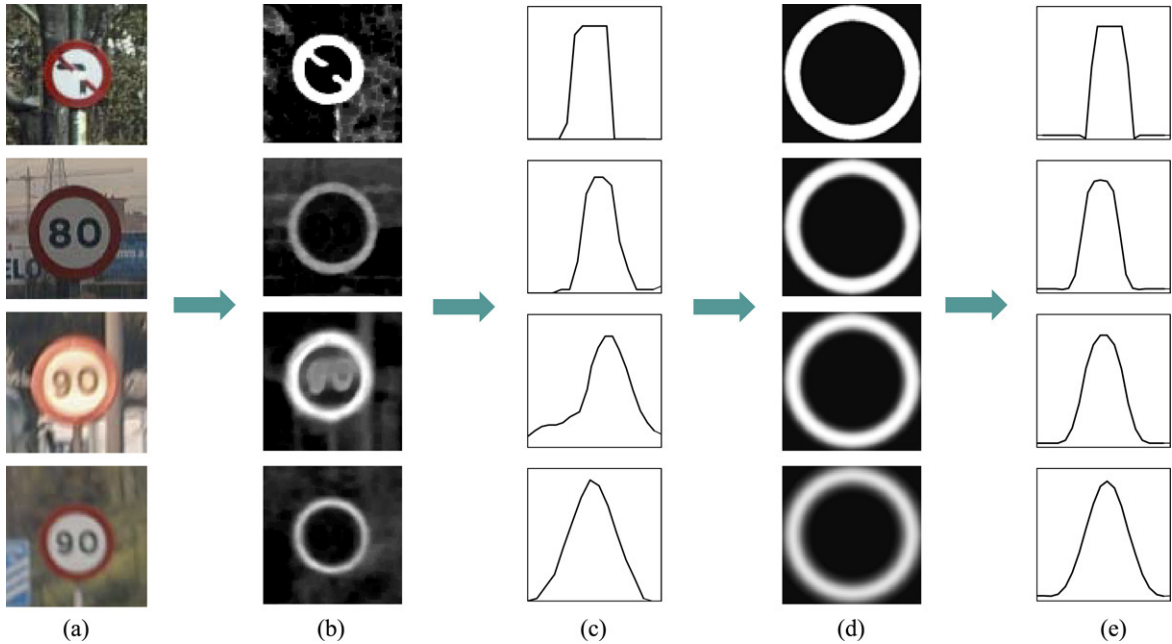


Fig. 13. Models for sign detection. (a) Images (b) colour enhanced (c) sign's border cross-section (d) models (e) model's border cross-section.

6.2. Traffic sign detection and classification

Traffic sign detection is performed through normalized correlation. Models of circular and triangular signs with different sizes are looked for in the image. As the same sign can give a high response to different models, they are passed in a certain order and previous detections in the same images are taken into account. Thus, the triangular models are correlated

first, as a circular sign can fit inside a triangular one and the models are used starting from the greatest to the smallest.

Besides that, the models do not have a sharp edge because it has been observed that the sign's appearance in the image has fuzzy edges. This is due to weather conditions, sign distance from the cameras and colour digitalisation. This has been translated to the models giving them a Gaussian form (Fig. 13).



Fig. 14. Two examples of the output of the system. The column beside the detected signs expresses the confidence value of the classification.

If there is a correlation value above a threshold the classification is performed. To do this, the region in the image is extracted, converted to a grey level image and normalized to another image of 32×32 pixels. The resulting image is a model that is correlated with another containing the whole set of possible signs (42 for circular, 43 for triangular, 11 for end-of-limit signs). If the correlation value is above a threshold the classification is valid and the information is used in conjunction with the vehicle's speed as was explained at the beginning of this section.

6.3. Results

The HSI colour space gives better results for 25% than the RGB, taking into account the detected signs and the false positives. The models with fuzzy edges give better results for 5% than the sharp ones. Fig. 14 shows the output of the system.

Two different audio warning messages can be emitted by the vehicles loudspeakers: the speed obtained through the GPS is inadequate for the recognized limit speed traffic sign; if it is a warning sign or yield sign and the vehicle speed is above a threshold which can be set by the driver.

7. Conclusions

In this article, the utility of applying Robotics and Artificial Intelligence techniques to driving safety, through the development of Advance Driver Assistance Systems, has been presented. Several modules, based on Computer vision, have been shown that allow the perception of road lanes, pedestrians, vehicles and traffic signs. This information can be useful for some ADAS like Line Keeping System, Adaptive Cruise Control, Pedestrian Protector, or Speed Supervisor. Finally, all the algorithms are tested in a research platform where they face real traffic situations and weather conditions.

Acknowledgment

This work was supported in part by the Spanish government under CICYT grant TRA2004-07441-C03-01.

References

- [1] E.D. Dickmanns, Vision for ground vehicles: History and prospects, *International Journal of Vehicle Autonomous Systems* 1 (2002) 1–44.
- [2] A. Broggi, M. Bertozzi, A. Fascioli, G. Conte, Automatic Vehicle Guidance: The Experience of the ARGO Vehicle, World Scientific, 1999.
- [3] D. Pomerleau, T. Jochem, Rapidly adapting machine vision for automated vehicle steering, *IEEE Expert* 11 (1996) 19–27.
- [4] U. Franke, D. Gavrilu, S. Gorzig, F. Lindner, F. Paetzold, C. Wohler, Autonomous driving approaches downtown, *IEEE Intelligent Systems* 13 (1998) 40–48.
- [5] U. Handmann, T. Kalinke, C. Tzomakas, M. Werner, An image processing system for driver assistance, *Image and Vision Computing* 18 (2000) 367–376.
- [6] U. Franke, D. Gavrilu, A. Gern, S. Gorzig, R. Janssen, F. Paetzol, C. Wohler, From door to door. Principles and applications of computer vision for driver assistant systems, in: L. Vlacic, F. Harashima, M. Parent (Eds.), *Intelligent Vehicles Technologies*, Butterworth Heinemann, Oxford, 2001, pp. 131–188.
- [7] L. Vlacic, M. Parent, F. Harashima, *Intelligent Vehicle Technologies: Theory and Applications*, Butterworth-Heinemann, 2001.
- [8] K. Kluge, S. Lakshmanan, A deformable-template approach to lane detection, in: *Proceedings of the Intelligent Vehicles '95 Symposium*, September 1995, pp. 54–59.
- [9] A. Broggi, Robust real-time lane and road detection in critical shadow conditions, in: *Proceedings IEEE International Symposium on Computer Vision*, Coral Gables, FL, November 1995, pp. 353–358.
- [10] J.M. Collado, C. Hilaro, A. de la Escalera, J.M. Armingol, Self-calibration of an on-board stereo-vision system for driver assistance systems, in: *IEEE Intelligent Vehicle Symposium*, Tokyo, Japan, June 2006, pp. 156–162.
- [11] D.M. Gavrilu, M. Kunert, U. Lages, A multi-sensor approach for the protection of vulnerable traffic participants—the PROTECTOR project, in: *IEEE Instrumentation and Measurement Technology Conference*, vol. 3, 2001, pp. 2044–2048.

- [12] A. Broggi, M. Bertozzi, A. Fascioli, M. Sechi, Shape-based pedestrian detection, in: IEEE Intelligent Vehicles Symposium, Dearborn, USA, October 2000, pp. 215–220.
- [13] D.M. Gavrila, V. Philomin, Real-time object detection for smart vehicle, in: Proc. of IEEE International Conference On Computer Vision, Kerkyra, Greece, 1999, pp. 87–93.
- [14] C. Papageorgiou, T. Evgeniou, T. Poggio, A trainable pedestrian detection system, in: Proc. of Intelligent Vehicles, Germany, 1998, pp. 241–246.
- [15] D.J. Williams, M. Shah, A fast algorithm for active contours and curvature estimation, CVGIP: Image Understanding 55 (1992) 14–26.
- [16] J.M. Collado, C. Hilario, A. de la Escalera, J.M. Armingol, Model based vehicle detection for intelligent vehicles, in: IEEE Intelligent Vehicles Symposium, Parma, Italy, 2004, pp. 572–577.
- [17] Z. Sun, G. Bebis, R. Miller, On-road vehicle detection using optical sensors: A review, in: IEEE Intelligent Transportation Systems Conference, Washington, DC, USA, 2004, pp. 585–590.
- [18] D.E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, Boston, 1989.
- [19] J. Torresen, J.W. Bakke, L. Sekanina, Efficient recognition of speed limit signs, in: Proc. of the 7th Int. IEEE Conf. on Intelligent Transportation Systems, Washington, DC, USA, 2004, pp. 652–656.
- [20] Z. Shuang-dong, Z. Yi, L. Xiao-feng, Detection for triangle traffic sign based on neural network, in: IEEE Int. Conf. on Vehicular Electronics and Safety, Xi'an, Shanxi, China, 2005, pp. 25–28.
- [21] C. Bahlmann, Y. Zhu, R. Visvanathan, M. Pellkofer, T. Koehler, A system for traffic sign detection, tracking, and recognition using color, shape, and motion information, in: Proc. of IEEE Intelligent Vehicles Symposium, Las Vegas, ND, USA, 2005, pp. 255–260.
- [22] A. Soetedjo, K. Yamada, Fast and robust traffic sign detection, in: IEEE Int. Conf. on Systems, Man and Cybernetics, Hawaii, USA, 2005, pp. 1341–1346.
- [23] N. Barnes, A. Zelinsky, Real-time radial symmetry for speed sign detection, in: IEEE Intelligent Vehicles Symposium, Parma, Italy, 2004, pp. 566–571.



José María Armingol is an Associate Professor of Industrial Engineering in the Department of Systems Engineering and Automation (Intelligent Systems Lab) at Universidad Carlos III de Madrid. He received the Degree in Automation and Electronics Engineering from the Universidad Politécnica de Madrid (Spain) in 1992 and the Ph.D. degree in industrial engineering from the Universidad Carlos III de Madrid (Spain) in 1997. His research interests are in mobile robotics, image processing and pattern recognition for intelligent

autonomous vehicles.



Arturo de la Escalera graduated from the Universidad Politécnica de Madrid (Madrid, Spain) in Automation and Electronics Engineering in 1989, where he also obtained his Ph.D. degree in Robotics in 1995. In 1993 he joined the Department of Automation of Universidad Carlos III de Madrid, (Madrid, Spain) where he is an Associate Professor from 1996. His current research interests include Advanced Robotics and Intelligent Transportation Systems with special emphasis on vision sensor systems and image data

processing methods for environment perception and real-time pattern recognition. He is member of the Editorial Board of the International Journal of Information and Communication Technology and the Open Transportation Journal.



Cristina Hilario graduated from the Universidad de Deusto in Computer Science in 2000. In 2001 she joined the Intelligent Systems Group of the department of System Engineering and Automation of the Universidad Carlos III de Madrid, where she is a Ph.D. Student. Her current research interest includes Intelligent Transportation Systems and Neuroscience.



Juan Manuel Collado graduated from the Universidad Carlos III de Madrid in Automation and Electronics Engineering in 2001. In 2002 he joined the Intelligent Systems Group of the department of System Engineering and Automation, where he is a Ph.D. Student. His current research interest includes Intelligent Transportation Systems and Ethics in Engineering Education.



Juan Pablo Carrasco received the Physics degree from the Universidad Autónoma de Madrid, Madrid, Spain, in 2001. Since 2005 he is a member of the Intelligent Systems Lab. Group of the Systems and Automation Engineering of the Universidad Carlos III de Madrid. His research interests include colour enhancement, pattern recognition and communications systems.



Marco Javier Flores received the B.S. degree in mathematics in 2001 from Escuela Politécnica Nacional (Ecuador) and the M.Sc. degree in statistics in 2004 from Pontificia Universidad Católica (Chile). He is a professor and researcher of computer vision and he is currently working in the Ph.D. thesis in Driver's drowsiness analysis, both with the Department of System and Automation, Universidad Carlos III de Madrid (Spain).



José Manuel Pastor received the Electronics and Automatic Control Engineering degree from the Polytechnic University of Madrid (UPM) in 1991 and a Ph.D. in Robotics and Artificial Intelligence in 1997. He has been Associate Professor since year 2000 in Univ. Carlos III of Madrid, and in 2005 Dr. Pastor moves to the Informatics Systems Department of the Castilla-La Mancha University where he is currently Associate Professor. His research interest includes intelligent transportation systems, and systems

simulation and optimization.



Francisco José Rodríguez is an Associate Professor of System Engineering and Automation at Carlos III University at Madrid, Spain, where he is the Head of the Intelligent Systems Lab since 2000. He received his Ph.D. degree in Artificial Intelligence and Robotics from the Polytechnic University of Madrid, Spain, in 1993. His current research interests are in robotics, online robots and intelligent systems.